Default Prior Choice for Bayesian Model Selection in Generalized Linear Models with Applications in Mortgage Default

by

Zachary Kramer

Department of Statistical Science and Economics
Duke University

Date: __________________________
Approved:

____
Merlise Clyde, Supervisor

____
Li Ma

____
Jia Li

Thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science in the Department of Statistical Science and Economics
in the Graduate School of Duke University
2014

Abstract

Default Prior Choice for Bayesian Model Selection in Generalized Linear Models with Applications in Mortgage Default

by

Zachary Kramer

Department of Statistical Science and Economics
Duke University

Date: ______________________
Approved:

____
Merlise Clyde, Supervisor

____
Li Ma

____
Jia Li

An abstract of a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Statistical Science and Economics
in the Graduate School of Duke University
2014

Abstract

The adoption of Zellner's g prior is a popular prior choice in Bayesian Model Averaging, although literature has shown that using a fixed g has undesirable properties. Mixtures of g priors have recently been proposed for Generalized linear models, extending results from the Gaussian linear model context. This paper will specifically look at the model selection problem as it applies to logistic regression. The effect of prior choice on both model selection and prediction using Bayesian Model Averaging is analyzed. This is done by testing a variety of model space and mixtures of g priors in a simulated data study as well illustrating their use in mortgage default data. This paper shows that the different mixtures of g priors tends to fall into one of two groups that have similar behavior. Additionally, priors in one of these groups, specifically the n/2, Beta Prime, and Robust mixtures of g priors, tend to outperform the other choices.

Contents

Abstract **iv**

List of Tables **vii**

List of Figures **x**

Acknowledgements **xi**

1 Introduction **1**

2 Bayesian Model Selection and Bayesian Model Averaging in Generalized Linear Models **3**

3 Prior Specification and Computation **5**

3.1 Priors for the Model Space 5

3.2 Zellner's g prior for α, β in GLMs 6

3.3 Priors for g and Computation of Bayes Factors 6

4 Simulated Data **12**

4.1 Description of the Simulated Data Study 12

4.2 Simulated Data Results: Model Selection 13

4.2.1 Larger Coefficients 13

4.2.2 Smaller Coefficients 16

4.3 Simulated Data Results: Prediction 20

4.3.1 Larger Coefficients 20

4.3.2 Smaller Coefficients 22

4.4 Simulated Data: Summary 24

5 Application to Mortgage Default 27

5.1 Data and Models 27

5.2 Mortgage Data Results: Model Selection 29

5.3 Mortgage Data Results: Prediction 33

6 Conclusion 42

Bibliography 43

List of Tables

4.1 Percent of Time True Model is Chosen: Larger Coefficients, Uniform Prior 14

4.2 Percent of Time Correct Model is Chosen: Larger Coefficients, Beta-Binomial(1,1) Prior 14

4.3 Avg Posterior Prob of True Model: Larger Coefficients, Uniform Prior 14

4.4 Avg Posterior Prob of True Model: Larger Coefficients, Beta-Binomial(1,1) Prior 15

4.5 Percent of Time Closest Model is Chosen: Larger Coefficients, Uniform Prior, Missing Predictor 15

4.6 Percent of Time Closest Model is Chosen: Larger Coefficients, Beta-Binomial(1,1) Prior, Missing Predictor 15

4.7 Avg Posterior Prob of Closest Model: Larger Coefficients, Uniform Prior, Missing Predictor 16

4.8 Avg Posterior Prob of Closest Model: Larger Coefficients, Beta-Binomial(1,1) Prior, Missing Predictor 16

4.9 Percent of Time True Model is Chosen: Smaller Coefficients, Uniform Prior 16

4.10 Percent of Time Correct Model is Chosen: Smaller Coefficients, Beta-Binomial(1,1) Prior 17

4.11 Avg Posterior Prob of True Model: Smaller Coefficients, Uniform Prior 17

4.12 Avg Posterior Prob of True Model: Smaller Coefficients, Beta-Binomial(1,1) Prior 17

4.13 Percent of Time Closest Model is Chosen: Smaller Coefficients, Uniform Prior, Missing Predictor 18

4.14 Percent of Time Closest Model is Chosen: Smaller Coefficients, Beta-Binomial(1,1) Prior, Missing Predictor 18

4.15 Avg Posterior Prob of Closest Model: Smaller Coefficients, Uniform Prior, Missing Predictor 19

4.16 Avg Posterior Prob of Closest Model: Smaller Coefficients, Beta-Binomial(1,1) Prior, Missing Predictor 19

4.17 Prediction Accuracy: Larger Coefficients, Uniform Prior 20

4.18 Prediction Accuracy: Larger Coefficients, Beta-Binomial(1,1) Prior . 20

4.19 Mean Root Error: Larger Coefficients, Uniform Prior 20

4.20 Mean Root Error: Larger Coefficients, Beta-Binomial(1,1) Prior . . . 21

4.21 Prediction Accuracy: Larger Coefficients, Uniform Prior, Missing Predictor 21

4.22 Prediction Accuracy: Larger Coefficients, Beta-Binomial(1,1) Prior, Missing Predictor 21

4.23 Mean Root Error: Larger Coefficients, Uniform Prior, Missing Predictor 22

4.24 Mean Root Error: Larger Coefficients, Beta-Binomial(1,1) Prior, Missing Predictor 22

4.25 Prediction Accuracy: Smaller Coefficients, Uniform Prior 22

4.26 Prediction Accuracy: Smaller Coefficients, Beta-Binomial(1,1) Prior . 23

4.27 Mean Root Error: Smaller Coefficients, Uniform Prior 23

4.28 Mean Root Error: Smaller Coefficients, Beta-Binomial(1,1) Prior . . . 24

4.29 Prediction Accuracy: Smaller Coefficients, Uniform Prior, Missing Predictor 24

4.30 Prediction Accuracy: Smaller Coefficients, Beta-Binomial(1,1) Prior, Missing Predictor 25

4.31 Mean Root Error: Smaller Coefficients, Uniform Prior, Missing Predictor 25

4.32 Mean Root Error: Smaller Coefficients, Beta-Binomial(1,1) Prior, Missing Predictor 26

5.1 Included Predictors by Model 37

5.2 Model Space Posterior: Uniform Prior 38
5.3 Marginal Inclusion Probabilities: Uniform Prior 38
5.4 Model Space Posterior: Beta-Binomial(1,1) Prior 39
5.5 Marginal Inclusion Probabilities: Beta-Binomial(1,1) 39
5.6 Grouped Predicted vs. Actual Default Rates: Uniform Prior 40
5.7 Total Predictions in Group: Uniform Prior 40
5.8 Grouped Predicted vs. Actual Default Rates: Beta-Binomial(1,1) Prior 40
5.9 Total Predictions in Group: Beta-Binomial(1,1) Prior 40
5.10 Avg Posterior β's: Uniform Prior 41
5.11 Avg Posterior β's: Beta-Binomial(1,1) Prior 41

List of Figures

3.1 Density of Prior Distributions for z 11

5.1 Effect of Credit Score and Number of Borrowers on η: n/2 Prior . . . 34

5.2 Effect of Credit Score and Number of Borrowers on η: Uniform Prior 35

Acknowledgements

Firstly, I would like to thank my advisor Merlise Clyde. She has been a great advisor, taking plenty of time to thoroughly answer any questions I have had while developing my thesis. I also had the honor of having Merlise as a professor where she introduced me to the concept of Bayesian Model Selection. Without her I never would of been able to complete this paper. Additionally, I would like to thank Yingbo Li, a former PhD student of Merlise's whose thesis formed the foundation for my work.

I would like to also thank my other committee members, Li Ma and Jia Li for taking the time out of their schedules to be on my thesis committee as well as being fantastic professors when I took classes with them. My thanks also goes out to the faculty of the Departments of Statistical Science and Economics at Duke University. I learned a lot from all of you during my time at Duke and feel I have been left in a great place to succeed in my future jobs and studies.

Lastly, I would like to thank my family for their support throughout my studies.

1

Introduction

A fully Bayesian analysis for model selection (BMS) (and/or model averaging (BMA)) requires putting a prior on the model space and all other model parameters. For generalized linear models (GLMs), a common choice of prior for α and β for Bayesian Model Averaging is Zellner's g prior. In order to avoid picking a fixed g for Zellner's g prior, which can have significant affect on the modeling and may be far from optimal, it is beneficial to put a prior on the parameter g. This has been studied extensively in papers such as Fernandez et al. (2001), Ley and Steel (2009) and (2011), Feldkircher and Zeugner (2009), and Liang et al. (2008).

I will look at the effect of prior choice on the model space and the effect of prior choice on g in BMS and BMA for logistic regression models. I will also compare the results of BMA to choosing a single model using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). I will test these differences in a simulated data study as well as using mortgage default data.

The rest of the paper is organized as follows: Section 2 is a general description of BMS/BMA, Section 3 describes the different priors that will be tested as well as the details of posterior computation, Section 4 contains the results from the simulated data study, and Section 5 contains the results from the mortgage default data.

2

Bayesian Model Selection and Bayesian Model Averaging in Generalized Linear Models

Say we have n data points with independent responses contained in vector Y and p potential predictors. We would like to specify a generalized linear model,

$$E(\mathbf{Y}) = \mathbf{1}\alpha + \mathbf{X}\boldsymbol{\beta}$$

with design matrix $\mathbf{X}$ containing some (or all) of the predictors and coefficients $\boldsymbol{\beta}$. Assuming we would only like to consider models with (or without) an intercept, there are 2^p potential models. The intercept term is separated from the other parameters as it is included in all considered models. Denote model $\mathcal{M}$ with design matrix $\mathbf{X}_{\mathcal{M}}$ containing columns $p_{\mathcal{M}}$ of the full design matrix $\mathbf{X}$ as

$$\mathcal{M}_j : \quad E(\mathbf{Y}) = \mathbf{1}\alpha_{\mathcal{M}} + \boldsymbol{X}_{\boldsymbol{\mathcal{M}}}\boldsymbol{\beta}_{\boldsymbol{\mathcal{M}}}$$

In order to implement a Bayesian solution to this problem, we need to implement priors on the model space, $\mathcal{M}$, and the parameters in each model, $\alpha_{\mathcal{M}}$ and $\boldsymbol{\beta}_{\boldsymbol{\mathcal{M}}}$. Once these priors are specified, we calculate the posterior probability of each model as follows:

$$p(\mathcal{M} \mid Y) = \frac{p(\mathcal{M}) \iiint p(\mathbf{Y} \mid \mathcal{M}, \alpha_{\mathcal{M}}, \boldsymbol{\beta}_{\mathcal{M}}) p(\alpha_{\mathcal{M}}, \boldsymbol{\beta}_{\mathcal{M}} \mid \mathcal{M}) \mathrm{d}\alpha_{\mathcal{M}} \mathrm{d}\boldsymbol{\beta}_{\mathcal{M}}}{\sum_{\mathcal{M}} p(\mathcal{M}) \iiint p(\mathbf{Y} \mid \mathcal{M}, \alpha_{\mathcal{M}}, \boldsymbol{\beta}_{\mathcal{M}}) p(\alpha_{\mathcal{M}}, \boldsymbol{\beta}_{\mathcal{M}} \mid \mathcal{M}) \mathrm{d}\alpha_{\mathcal{M}} \mathrm{d}\boldsymbol{\beta}_{\mathcal{M}}}$$

Once these posterior probabilities are calculated, we can select the model with the highest posterior probability or implement Bayesian Model Averaging (BMA). In BMA we use all of our considered models and set β's using an average of the expected value of β of each individual model weighted by posterior probabilities of each model. So,

$$\hat{\beta}_j = \sum_{\mathcal{M}} E[\beta_j \mid \mathcal{M}] p(\mathcal{M} \mid \mathbf{Y})$$

where $E[\beta_j \mid \mathcal{M}] = 0$ if predictor j is not included in model $\mathcal{M}$.

Now, we need priors $p(\mathcal{M})$ on the model space and priors on α and $\boldsymbol{\beta} \mid \mathcal{M}$ as well as ability to calculate integrals and explore the space of models. Use of Zellner's g prior avoids the need of Markov Chain Monte Carlo (MCMC) to integrate out β and thus reduces Monte Carlo variation.

3

Prior Specification and Computation

3.1 Priors for the Model Space

Two priors for the model space will be considered in the following analyses. The Uniform prior and the Beta-Binomial(1,1) prior.

In the Uniform prior, each model is assigned the same probability a priori. If there are p covariates in the full model, the a priori expectation of the model size is $\frac{p}{2}$ with a standard deviation of $\frac{\sqrt{p}}{2}$ (Scott, Berger 2010). So, for large p, the a priori distribution of model size is very concentrated around half the variables being included.

If we instead say that a priori each variable has the same probability, q, of being included, we have:

$$p(\mathcal{M} \mid q) = q^{p_{\mathcal{M}}}(1-q)^{p-p_{\mathcal{M}}}$$

and that $p_{\mathcal{M}}$ has a Binomial distribution, $Bin(p, q)$.

The Uniform prior is equivalent to the above representation, assigning a constant value of 0.5 to q. Using any constant value for q will not allow you to control for issues regarding multiple testing and may result in an increasing number of false positive as p increases (Scott, Berger 2010).

The other prior that will be considered, the Beta-Binomial(1,1) prior, corrects for this issue by placing a Beta(1,1) prior on q instead of setting it to a fixed value. This prior results in a multiplicity penalty (Scott, Berger 2010). After marginalizing out q, the a priori probability of model $\mathcal{M}$ is:

$$p(\mathcal{M}) = \frac{1}{p}\binom{p}{p_{\mathcal{M}}}^{-1}$$

where p is the number of predictors in the full model.

3.2 Zellner's g prior for α, β in GLMs

We assign priors to the intercept, $\alpha_{\mathcal{M}}$ and $\beta_{\mathcal{M}}$'s as follows:

$$\boldsymbol{\beta_{\mathcal{M}}} \mid g, \mathcal{M} \sim N(0, g * \boldsymbol{\mathcal{I}_n}(\hat{\boldsymbol{\beta_{\mathcal{M}}}})^{-\mathbf{1}})$$

$$\alpha_{\mathcal{M}} \mid \mathcal{M} \sim N(0, nc)$$

With positive parameter g and non-negative parameter c. Under this prior, it can be shown (Li, 2013) that the posterior distributions converge as follows:

$$\alpha_{\mathcal{M}} \mid \mathbf{Y}, \mathcal{M} \to N\left(\frac{\mathcal{I}_n(\hat{\alpha}_{\mathcal{M}})}{\mathcal{I}_n(\hat{\alpha}_{\mathcal{M}}) + \frac{1}{nc}}\hat{\alpha}_{\mathcal{M}}, \frac{1}{\mathcal{I}_n(\hat{\alpha}_{\mathcal{M}}) + \frac{1}{nc}}\right)$$

$$\boldsymbol{\beta_{\mathcal{M}}} \mid g, \mathbf{Y}, \mathcal{M} \to N\left(\frac{g}{1+g}\hat{\boldsymbol{\beta}}_{\boldsymbol{\mathcal{M}}}, \frac{g}{1+g}\boldsymbol{\mathcal{I}_n}(\hat{\boldsymbol{\beta}}_{\boldsymbol{\mathcal{M}}})^{-\mathbf{1}}\right)$$

and unless g adapts with the sample, inference under fixed g maybe inconsistent. Li suggests priors on g to address this.

3.3 Priors for g and Computation of Bayes Factors

First, define z as the shrinkage factor where:

$$z = \frac{g}{1+g}$$

In the following analysis, I will use the following family of priors for z, known as the The Confluent Hypergeometric Distribution (Gordy, 1998):

$$p(z \mid a, b, s) = \frac{z^{a-1}(1-z)^{b-1}exp[-sz]}{B(a,b)_1F_1(a, a+b, -s)}, 0 < z < 1$$

Refer to this as the CH(a,b,s) distribution. This becomes the Beta(a,b) distribution when s=0. According to Li, 2008, the resulting posterior distribution of z under the prior $z \sim CH(\frac{a}{2}, \frac{b}{2}, -\frac{s}{2})$ is

$$z \mid \mathbf{Y}, \mathcal{M} \sim CH(\frac{a}{2}, \frac{b+p_{\mathcal{M}}}{2}, -\frac{s+Q_{\mathcal{M}}}{2})$$

where

$$Q_{\mathcal{M}} = [\hat{\boldsymbol{\beta}}^{\boldsymbol{T}}_{\boldsymbol{\mathcal{M}}} \boldsymbol{X}^{\boldsymbol{T}}_{\boldsymbol{\mathcal{M}}}] \boldsymbol{\mathcal{I}_n}(\hat{\boldsymbol{\eta}}_{\boldsymbol{\mathcal{M}}})[\boldsymbol{X}_{\boldsymbol{\mathcal{M}}} \hat{\boldsymbol{\beta}}_{\boldsymbol{\mathcal{M}}}]$$

similar to RSS in linear models. Under the g prior, it can be shown using Laplace Approximation (Li, 2013) that the marginal likelihood can be approximated by

$$p(\mathbf{Y} \mid g, \mathcal{M}) = f_{\mathcal{M}}(\mathbf{Y} \mid \hat{\alpha}_{\mathcal{M}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{\mathcal{M}}})[1 + \mathcal{I}_n(\hat{\alpha}_{\mathcal{M}})nc]^{-\frac{1}{2}} e^{-\frac{\mathcal{I}_n(\hat{\alpha}_{\mathcal{M}})\hat{\alpha}^2_{\mathcal{M}}}{2(\mathcal{I}_n(\hat{\alpha}_{\mathcal{M}})^{nc+1})}} (1+g)^{-\frac{p_{\mathcal{M}}}{2}} e^{\frac{-Q_{\mathcal{M}}}{2(1+g)}}$$

and then the Bayes Factor for comparing $\mathcal{M}_1$ to $\mathcal{M}_2$ can be approximated using a Laplace Approximation (Li, 2013) by:

$$BF_{\mathcal{M}_2:\mathcal{M}_1} = \Lambda_{\mathcal{M}_2:\mathcal{M}_1} \Omega_{\mathcal{M}_2:\mathcal{M}_1}$$

where

$$\Lambda_{\mathcal{M}_2:\mathcal{M}_1} = \frac{f_{\mathcal{M}_2}(\mathbf{Y} \mid \hat{\alpha}_{\mathcal{M}_2}, \hat{\boldsymbol{\beta}}_{\boldsymbol{\mathcal{M}_2}})}{f_{\mathcal{M}_1}(\mathbf{Y} \mid \hat{\alpha}_{\mathcal{M}_1}, \hat{\boldsymbol{\beta}}_{\boldsymbol{\mathcal{M}_1}})} \left(\frac{1 + \mathcal{I}_n(\hat{\alpha}_{\mathcal{M}_2})nc}{1 + \mathcal{I}_n(\hat{\alpha}_{\mathcal{M}_1})nc} \right)^{-1/2} *$$

$$exp\left(-\frac{1}{2} \left(\frac{\mathcal{I}_n(\hat{\alpha}_{\mathcal{M}_2})\hat{\alpha}^2_{\mathcal{M}_2}}{\mathcal{I}_n(\hat{\alpha}_{\mathcal{M}_2})^{nc+1}} - \frac{\mathcal{I}_n(\hat{\alpha}_{\mathcal{M}_1})\hat{\alpha}^2_{\mathcal{M}_1}}{\mathcal{I}_n(\hat{\alpha}_{\mathcal{M}_1})^{nc+1}} \right) \right)$$

$$\Omega_{\mathcal{M}_1:\mathcal{M}_2} = \frac{B(\frac{b+p_{\mathcal{M}_2}}{2}, \frac{a}{2})_1F_1(\frac{b+p_{\mathcal{M}_2}}{2}, \frac{a+b+p_{\mathcal{M}_2}}{2}, -\frac{s+Q_{\mathcal{M}_2}}{2})}{B(\frac{b+p_{\mathcal{M}_1}}{2}, \frac{a}{2})_1F_1(\frac{b+p_{\mathcal{M}_1}}{2}, \frac{a+b+p_{\mathcal{M}_1}}{2}, -\frac{s+Q_{\mathcal{M}_1}}{2})}$$

The priors for g that will be considered in the following analyses are the Uniform prior (Wang and George, 2007), the Hyper-g prior (Liang et al., 2008), Jeffrey's prior (Celeux et al., 2012), the Beta Prime (Beta) prior (Maruyama and George, 2011) the n/2 prior (Li, 2013), the Robust prior (Bayarri et al., 2012), and Local Empirical Bayes (EB) (Hansen and Yu, 2003). The Uniform, Hyper-g, Jeffrey's, Beta Prime, and n/2 priors are part of the CH(a,b,s) family of priors. I will also compare these to AIC and BIC. Below is a brief description of each of the above:

Uniform Prior: Use the Uniform(0,1) distribution as the prior on z.

Hyper-g: (Liang et al., 2008) introduces the Hyper-g family of priors for g for normal linear models, defined as

$$p(g) = \frac{a-2}{2}(1+g)^{-a/2}, g > 0$$

Equivalently as a prior on z:

$$z \sim Beta(1, \frac{a}{2} - 1)$$

with a recommended value of $2 < a < 4$ for linear models. In this paper $a = 3$ will be used.

Jeffrey's: (Celeux et al., 2012) suggests using Jeffreys prior as a non-informative prior on g for normal linear models where:

$$p(g) \propto \frac{1}{1+g}$$

Beta Prime: (Maruyama and George, 2011) describe the Beta Prime family of priors on g for normal linear models as:

$$p(g) = \frac{g^b(1+g)^{-a-b-2}}{B(a+1,b+1)}, g > 0$$

Equivalently, as a prior on $\frac{1}{1+g}$,

$$\frac{1}{1+g} \sim Beta(a+1, b+1)$$

In this paper $a = -0.75$ and $b = \frac{n-p_{\mathcal{M}}}{2} - 1.75$ will be used.

n/2: (Li, 2013) suggests the following prior for generalized linear models:

$$g \sim CH(n/2, 0.5, 0)$$

Robust: (Bayarri et al., 2012) discusses the Robust family of priors on g for normal linear models:

$$p(g) = a[\rho(b+n)]^a(g+b)^{-(a+1)}, g > \rho(b+n) - b$$

or as shown in (Li, 2013) after transforming g to $u = 1/(1+g)$:

$$p(u) = a[\rho(b+n)]^a \frac{u^{a-1}}{[1+(b-1)u]^{a+1}}, 0 < u < \frac{1}{\rho(b+n)+(1-b)}$$

In this paper, I will use the values recommended in Bayarri which are $a = 0.5$, $b = 1$, and $\rho = \frac{1}{1+p_{\mathcal{M}}}$

Local EB: (Hansen and Yu, 2003) describe an Empirical Bayesian procedure for choosing g in generalized linear models where:

$$\hat{g}_{\mathcal{M}} = max(\frac{Q}{p_{\mathcal{M}}} - 1, 0)$$

AIC and BIC: Model selection criterion where the chosen model has the lowest AIC and BIC values respectively where:

$$AIC = 2p_{\mathcal{M}} - 2log(L(y))$$

$$BIC = p_{\mathcal{M}}log(n) - 2log(L(Y))$$

where $L(Y)$ is the likelihood of Y evaluated at the MLEs.

Figure 3.1 shows the prior distributions for z when $n = 250$ and $p_{\mathcal{M}} = 6$ where these are needed. This corresponds to the values for the full model in the simulated data study. The Robust prior is not shown as a priori it only has positive density close to one.

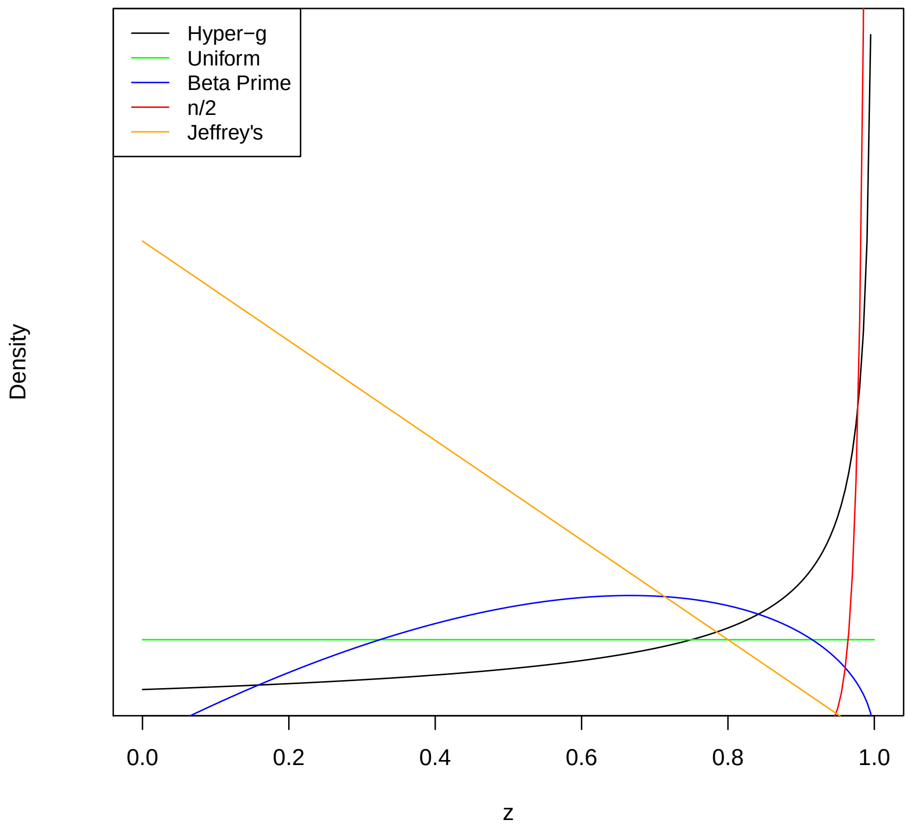

FIGURE 3.1: Density of Prior Distributions for z

4

Simulated Data

4.1 Description of the Simulated Data Study

In order to see the effects of prior choice, I generated data under 12 different models.

Larger Coefficients

$$m_1 : logit(\pi) = 2 + 2 * x_1$$

$$m_2 : logit(\pi) = 2 + 2 * x_1 - 2.5 * x_2 + 3 * x_3$$

$$m_3 : logit(\pi) = 2 + 2 * x_1 - 2.5 * x_2 + 3 * x_3 - 2 * x_4 + 2 * x_5 - 3 * x_6$$

$$m_4 : logit(\pi) = 2 + 2 * x_1 + 2.5 * x_7$$

$$m_5 : logit(\pi) = 2 + 2 * x_1 - 2.5 * x_2 + 3 * x_3 + 2.5 * x_7$$

$$m_6 : logit(\pi) = 2 + 2 * x_1 - 2.5 * x_2 + 3 * x_3 - 2 * x_4 + 2 * x_5 - 3 * x_6 + 2.5 * x_7$$

Smaller Coefficients

$$m_7 : logit(\pi) = 1 + 1 * x_1$$

$$m_8 : logit(\pi) = 1 + 1 * x_1 - 1.25 * x_2 + 1.5 * x_3$$

$$m_9 : logit(\pi) = 1 + 1 * x_1 - 1.25 * x_2 + 1.5 * x_3 - 1 * x_4 + 1 * x_5 - 1.5 * x_6$$

$$m_{10} : logit(\pi) = 1 + 1 * x_{+}2.5 * x_7 1$$

$$m_{11} : logit(\pi) = 1 + 1 * x_1 - 1.25 * x_2 + 1.5 * x_3 + 2.5 * x_7$$

$$m_{12} : logit(\pi) = 1 + 1 * x_1 - 1.25 * x_2 + 1.5 * x_3 - 1 * x_4 + 1 * x_5 - 1.5 * x_6 + 2.5 * x_7$$

Models 4-6 and 10-12 are the same as models 1-3 and 7-9 with the addition of a predictor, x_7 that will not be considered in the model selection process in order to simulate a situation in which the response also relies significantly on a covariate not included in the data set. These are the data sets used in the "Missing Predictor" tables that follow.

All x_i's (including the "missing predictor") are standard normal variables. 500 datasets of 250 points were generated. Under each of the 12 models, I will look at how often the actual model, or in the case of the missing predictor models the closest model, is chosen as well as the resulting posterior probability of this model. For prediction, I will consider two criteria; prediction accuracy on a 0-1 basis and how close the predicted π is to the true π.

Two priors will be considered on the model space. These are the Uniform prior and the Beta-Binomial(1,1) prior. Prediction results will be analyzed under Bayesian Model Averaging for all of the described priors for g in Section 3.3. Specifically, these are the n/2, Uniform, Hyper-g, Jeffrey's, Beta Prime, Local EB, and Robust priors. I will also get results under AIC and BIC in which prediction will just be made using the top model.

4.2 Simulated Data Results: Model Selection

4.2.1 Larger Coefficients

Under the models with larger coefficients, Tables 4.1 and 4.2 show the percentage of the time the true model was chosen under each g prior and model space prior. The model described as chosen is the one with the highest posterior likelihood.

Table 4.1: Percent of Time True Model is Chosen: Larger Coefficients, Uniform Prior

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_1	0.87	0.61	0.69	0.80	**0.91**	0.41	**0.91**	0.70	0.85
m_2	0.87	0.59	0.65	0.70	0.89	0.53	**0.94**	0.65	0.89
m_3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 4.2: Percent of Time Correct Model is Chosen: Larger Coefficients, Beta-Binomial(1,1) Prior

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_1	0.95	0.79	0.85	0.91	**0.97**	0.41	0.91	0.85	0.94
m_2	0.77	0.10	0.22	0.37	0.84	0.53	**0.94**	0.23	0.84
m_3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

All models and prior combinations chose the correct model every time when the true model was the full model. The Beta-Binomial(1,1) prior performed better when m_1 was the true model and the Uniform prior performed better when m_2 was the true model. In general, BIC and the n/2, Beta Prime, and Robust priors chose the correct model significantly more often than AIC and the Uniform, Hyper-g, Jeffrey's, and Local EB priors. The latter had significant trouble in choosing the correct model under the Beta-Binomial(1,1) prior.

Tables 4.3 and 4.4 show the average posterior likelihoods of the true model under each prior combination.

Table 4.3: Avg Posterior Prob of True Model: Larger Coefficients, Uniform Prior

	n/2	Uniform	Hyp-g	Jeffreys	Beta	EB	Robust
m_1	**0.45**	0.16	0.21	0.31	0.53	0.21	0.42
m_2	0.50	0.26	0.30	0.34	**0.57**	0.30	0.56
m_3	1.00	1.00	1.00	1.00	1.00	1.00	1.00

When the true model was the full model, the posterior probability of the full model was one under all prior combinations. As expected, the posteriors probabilities of the true model are higher for m_2 when the Uniform model space prior is used and for m_1 when the Beta-Binomial(1,1) prior is used. The Beta Prime prior results in

Table 4.4: Avg Posterior Prob of True Model: Larger Coefficients, Beta-Binomial(1,1) Prior

	n/2	Uniform	Hyp-g	Jeffreys	BetaPrime	EB	Robust
m_1	0.66	0.30	0.37	0.51	**0.73**	0.38	0.63
m_2	0.35	0.12	0.14	0.17	**0.43**	0.14	0.42
m_3	1.00	1.00	1.00	1.00	1.00	1.00	1.00

the highest posterior probability for both m_1 and m_2 within each model space prior selection. The Uniform prior results in the lowest posterior probability for both m1 and m2 within each model space prior selection. The n/2, BetaPrime, and Robust priors outperform the rest of the priors just as they did in the previous analysis.

I will now repeat this analysis for the larger coefficient models with a missing predictor.

Table 4.5: Percent of Time Closest Model is Chosen: Larger Coefficients, Uniform Prior, Missing Predictor

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_4	0.88	0.56	0.66	0.79	**0.91**	0.38	**0.91**	0.65	0.85
m_5	0.87	0.67	0.71	0.76	0.90	0.57	**0.93**	0.71	0.90
m_6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 4.6: Percent of Time Closest Model is Chosen: Larger Coefficients, Beta-Binomial(1,1) Prior, Missing Predictor

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_4	0.95	0.75	0.83	0.91	**0.96**	0.38	0.91	0.83	0.94
m_5	0.78	0.24	0.37	0.52	0.86	0.57	**0.93**	0.38	0.84
m_6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Tables 4.5 and 4.6 show that under both model space priors, adding the missing predictor does not have a large affect on the percent of the time the correct model is chosen. The one exception is that when the mid sized model is the correct model, the probability of the correct model being chosen under the Uniform, Hyper-g, Jeffrey's, and Local EB priors increases although these still perform poorly compared to the

other choices. This change is relatively small under the Uniform model space prior and much greater under the Beta-Binomial(1,1) prior.

Table 4.7: Avg Posterior Prob of Closest Model: Larger Coefficients, Uniform Prior, Missing Predictor

	n/2	Uniform	Hyp-g	Jeffreys	Beta	EB	Robust
m_4	0.44	0.13	0.18	0.27	**0.53**	0.18	0.42
m_5	0.51	0.30	0.33	0.37	**0.58**	0.33	0.57
m_6	0.99	**1.00**	**1.00**	**1.00**	0.99	**1.00**	0.99

Table 4.8: Avg Posterior Prob of Closest Model: Larger Coefficients, Beta-Binomial(1,1) Prior, Missing Predictor

	n/2	Uniform	Hyp-g	Jeffreys	Beta	EB	Robust
m_4	0.65	0.25	0.32	0.46	**0.72**	0.33	0.63
m_5	0.36	0.14	0.17	0.21	**0.44**	0.17	**0.44**
m_6	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Tables 4.7 and 4.8 show that when the missing predictor was added, the average posterior probability of the correct model decreased when the correct model was the small model and increased when the correct model was medium sized model regardless of prior choice on model space or g. This change was significantly larger for the Uniform, Hyper-g, Jeffrey's and Local EB priors.

4.2.2 Smaller Coefficients

In this analysis, coefficients were decreased to half of their previous values, except the coefficient on the missing predictor which was held at 2.5.

Table 4.9: Percent of Time True Model is Chosen: Smaller Coefficients, Uniform Prior

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_7	0.86	0.57	0.64	0.79	**0.89**	0.45	**0.89**	0.64	0.84
m_8	0.85	0.69	0.73	0.77	0.89	0.60	**0.93**	0.73	0.88
m_9	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.99	**1.00**	**1.00**

Table 4.10: Percent of Time Correct Model is Chosen: Smaller Coefficients, Beta-Binomial(1,1) Prior

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_7	0.94	0.75	0.81	0.88	**0.95**	0.45	0.89	0.81	0.94
m_8	0.79	0.21	0.36	0.53	0.85	0.60	**0.93**	0.36	0.84
m_9	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.99	**1.00**	**1.00**

Tables 4.9 and 4.10 show the percent of the time the correct model is chosen when the data is generated under the models with smaller coefficients without a missing predictor. Similar results hold here as to that under the larger coefficients. The n/2, Beta Prime and Robust priors and BIC significantly outperform the other priors and AIC. Also, the Beta-Binomial(1,1) prior is still best for choosing the small model correctly while the Uniform prior is better for choosing mid sized model correctly. When compared to the percent of time the correct model was chosen under the larger coefficients, the differences for the n/2, Beta Prime and Robust priors and BIC were very small. For the other priors, the probability of selecting the true model decreased when it was the small model while the probability of selecting the true model increased when it was the mid sized model.

Table 4.11: Avg Posterior Prob of True Model: Smaller Coefficients, Uniform Prior

	n/2	Uniform	Hyp-g	Jeffreys	Beta	EB	Robust
m_7	0.45	0.13	0.17	0.27	**0.53**	0.18	0.42
m_8	0.51	0.30	0.33	0.37	**0.58**	0.33	0.57
m_9	0.98	**0.99**	**0.99**	**0.99**	0.98	**0.99**	0.98

Table 4.12: Avg Posterior Prob of True Model: Smaller Coefficients, Beta-Binomial(1,1) Prior

	n/2	Uniform	Hyp-g	Jeffreys	Beta	EB	Robust
m_7	0.65	0.25	0.32	0.45	**0.72**	0.32	0.63
m_8	0.36	0.14	0.17	0.21	**0.44**	0.17	**0.44**
m_9	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Tables 4.11 and 4.12 show the average posterior probabilities of the true models under each prior choice. Surprisingly, they are nearly identical to their larger coeffi-

cient counterparts for the n/2, Beta Prime and Robust priors. So, the Beta Prime prior still outperforms the other priors. For the Uniform, Hyper-g, Jeffrey's, and Local EB priors, the posterior probability of the small model is slightly smaller and the posterior probability of the mid sized model is slightly larger than that under the larger coefficients.

Table 4.13: Percent of Time Closest Model is Chosen: Smaller Coefficients, Uniform Prior, Missing Predictor

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_{10}	0.81	0.47	0.56	0.69	0.82	0.43	**0.85**	0.54	0.84
m_{11}	0.81	0.63	0.68	0.73	0.84	0.56	**0.86**	0.68	0.85
m_{12}	0.83	0.94	0.93	0.91	0.79	**0.95**	0.65	0.93	0.77

Table 4.14: Percent of Time Closest Model is Chosen: Smaller Coefficients, Beta-Binomial(1,1) Prior, Missing Predictor

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_{10}	0.75	0.40	0.59	0.78	0.74	0.43	0.85	0.56	**0.93**
m_{11}	0.74	0.04	0.15	0.32	0.76	0.56	**0.86**	0.15	0.78
m_{12}	**1.00**	**1.00**	**1.00**	**1.00**	0.99	0.95	0.65	**1.00**	0.99

Tables 4.13 and 4.14 show the percentage of the time the correct model was selected for the models with the smaller coefficients and missing predictor. Generally, the percentage of the time the correct model was chosen across nearly all models and priors decreased. Under the Uniform model space prior, all priors start having trouble choosing the large model, with BIC and the n/2, BetaPrime, and Robust priors performing worse than the other priors for the first time under any scenario. They still significantly outperform the other priors in choosing the small and mid sized models though. For these models, the Robust prior nearly matches its numbers from that under the scenario without a missing predictor and does the best in choosing the small and mid sized models. Under the Beta-Binomial prior, the large model is again correctly chosen almost all the time. For the small and mid sized models, all probabilities of choosing the correct model decreased from that without the missing

predictor, but the decreases under the robust prior were small and the Robust prior performed the best.

Table 4.15: Avg Posterior Prob of Closest Model: Smaller Coefficients, Uniform Prior, Missing Predictor

	n/2	Uniform	Hyp-g	Jeffreys	Beta	EB	Robust
m_{10}	0.39	0.07	0.10	0.16	**0.47**	0.10	0.40
m_{11}	0.46	0.25	0.28	0.32	**0.51**	0.28	**0.51**
m_{12}	0.71	**0.80**	**0.80**	0.78	0.67	**0.80**	0.65

Table 4.16: Avg Posterior Prob of Closest Model: Smaller Coefficients, Beta-Binomial(1,1) Prior, Missing Predictor

	n/2	Uniform	Hyp-g	Jeffreys	Beta	EB	Robust
m_{10}	0.52	0.13	0.18	0.29	0.56	0.18	**0.60**
m_{11}	0.32	0.11	0.13	0.16	**0.39**	0.13	**0.39**
m_{12}	0.91	**0.95**	**0.95**	0.94	0.89	**0.95**	0.88

Tables 4.15 and 4.16 show the average posterior probabilities of the closest models. The effect of adding the missing predictor on the average posterior probabilities is the same as that on the percent of time the correct model is chosen. Under the Uniform model space prior especially, the posterior likelihood of the true model is much lower when the large model is correct. Also, the posterior probability of the correct model under all prior choices is lower. When the full model is the true model, the Uniform, Hyper-g, Jeffrey's and Local EB priors put the highest posterior likelihood on the closest model, but when the small and mid sized are the correct models their posterior likelihood of the closest model is very low.

4.3 Simulated Data Results: Prediction

For prediction, I generated a 500 point data set under each of the previous 12 models. I then tested the prediction accuracy under each of the 500 training data sets two ways. First, I tested the prediction accuracy by predicting 1 if $\hat{\pi} > 0.5$ and zero otherwise. Then, since we know the π_i's that generated the data, I calculated the mean root error (MRE) under each model and prior combination as

$$MRE = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \sqrt{|\hat{\pi}_i - \pi_i|}$$

4.3.1 Larger Coefficients

Table 4.17: Prediction Accuracy: Larger Coefficients, Uniform Prior

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_1	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84
m_2	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91
m_3	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89

Table 4.18: Prediction Accuracy: Larger Coefficients, Beta-Binomial(1,1) Prior

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_1	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84
m_2	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91
m_3	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89

Table 4.19: Mean Root Error: Larger Coefficients, Uniform Prior

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_1	**0.14**	0.16	0.15	0.15	**0.14**	0.16	**0.14**	0.15	**0.14**
m_2	**0.13**	0.14	0.14	0.14	**0.13**	**0.13**	**0.13**	0.14	**0.13**
m_3	0.14	0.16	0.16	0.15	0.14	**0.13**	**0.13**	0.15	0.14

Tables 4.17-4.20 show the prediction accuracy and MRE across the models with the larger coefficients and no missing predictor. For the prediction accuracy, choice of prior makes essentially no difference. This seems likely as the prediction for a

Table 4.20: Mean Root Error: Larger Coefficients, Beta-Binomial(1,1) Prior

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_1	**0.14**	0.16	0.15	0.15	**0.14**	0.16	**0.14**	0.15	**0.14**
m_2	**0.13**	0.15	0.15	0.14	**0.13**	**0.13**	**0.13**	0.14	**0.13**
m_3	0.14	0.16	0.16	0.15	0.14	**0.13**	**0.13**	0.15	0.14

given data point will only change if the change of prior causes the prediction of π_i to cross the 0.5 boundary. The prediction accuracy is the most accurate when the true model is m_2 and least accurate when the true model is m_1. This is a little surprising as the simulated data when m_1 is the true model has an average value of π farthest from 0.5, so you would expect it to be the easiest to predict. The average value of π under m_1 is 0.74, m_2 is 0.66, and m_3 is 0.63. Similar to the model selection results, MRE follows the same pattern in that the Beta-Binomial(1,1) prior performs better than the Uniform prior when m_1 is the true model and vice versa when m_2 is the true model. The results are identical under m_3 as the correct model is chosen essentially every time with close to a 100% posterior probability under all prior choices. Also, similar to the model selection results, BIC and the n/2, Beta Prime, and Robust priors outperform the other choices.

Table 4.21: Prediction Accuracy: Larger Coefficients, Uniform Prior, Missing Predictor

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_4	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76
m_5	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83
m_6	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85

Table 4.22: Prediction Accuracy: Larger Coefficients, Beta-Binomial(1,1) Prior, Missing Predictor

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_4	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76
m_5	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83
m_6	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85

Table 4.23: Mean Root Error: Larger Coefficients, Uniform Prior, Missing Predictor

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_4	**0.42**	0.43	0.43	0.43	**0.42**	**0.42**	**0.42**	0.43	**0.42**
m_5	0.34	0.35	0.35	0.35	0.34	**0.33**	**0.33**	0.34	0.34
m_6	0.31	0.33	0.32	0.32	0.31	**0.29**	**0.29**	0.32	0.31

Table 4.24: Mean Root Error: Larger Coefficients, Beta-Binomial(1,1) Prior, Missing Predictor

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_4	**0.42**	0.43	0.43	0.43	**0.42**	**0.42**	**0.42**	0.43	**0.42**
m_5	0.34	0.35	0.35	0.35	0.34	**0.33**	**0.33**	0.35	0.34
m_6	0.31	0.33	0.32	0.32	0.31	**0.29**	**0.29**	0.32	0.31

Tables 4.21-4.24 show the prediction accuracy and MRE across the models with the larger coefficients and a missing predictor. For prediction accuracy, choice of prior still makes essentially no difference. The missing predictor causes a drop in prediction accuracy of about 0.08 when the small and mid sized are the closest models and a drop of about 0.04 when the large model is the closest model. As one would expect, the MRE significantly increases with the inclusion of the missing predictor. Both model space prior choice and g prior choice do not affect the MRE very much. In general, the bigger the closest model was, the lower the affect on the MRE by adding the missing predictor.

4.3.2 Smaller Coefficients

Table 4.25: Prediction Accuracy: Smaller Coefficients, Uniform Prior

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_7	**0.73**	0.72	0.72	**0.73**	**0.73**	0.72	**0.73**	0.72	**0.73**
m_8	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83
m_9	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81

Tables 4.25-4.28 show the prediction accuracy and MRE for the various prior choices under the models with the smaller coefficients and no missing predictor. Similar to the models with the larger coefficients, prediction accuracy is essentially

Table 4.26: Prediction Accuracy: Smaller Coefficients, Beta-Binomial(1,1) Prior

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_7	**0.73**	0.72	**0.73**	**0.73**	**0.73**	0.72	**0.73**	0.72	**0.73**
m_8	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83
m_9	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81

Table 4.27: Mean Root Error: Smaller Coefficients, Uniform Prior

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_7	0.32	0.34	0.33	0.33	0.32	0.32	**0.31**	0.33	0.32
m_8	0.29	0.31	0.30	0.30	0.29	**0.27**	**0.27**	0.30	0.29
m_9	0.27	0.29	0.28	0.28	0.26	**0.25**	**0.25**	0.28	0.26

the same under all prior choices. Compared to the larger coefficients, prediction accuracy decreased under all prior choices. Some of this difference can be attributed to the fact that the average π under all models is slightly closer to 0.5 with the smaller coefficients. The MRE under all models and all priors essentially doubled from that under the larger coefficients. The differences between MRE under the different prior choices is nearly identical as well.

Tables 4.29-4.32 show the prediction accuracy and MRE for the various prior choices under the models with the smaller coefficients and a missing predictor. Prior choice still has little affect on prediction accuracy and the affect of adding the missing predictor on prediction accuracy is the same as that with the larger coefficients, but to a larger extent. Prior choice also has little affect on the MRE. Adding the missing predictor does cause an increase in MRE for all prior choices, but doesn't quite double it as with the larger coefficients. The MRE follows the same pattern as in the larger coefficients with a missing predictor in that as the model size of the closest model increases, the MRE decreases.

Table 4.28: Mean Root Error: Smaller Coefficients, Beta-Binomial(1,1) Prior

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_7	0.32	0.34	0.33	0.33	0.32	0.32	**0.31**	0.33	0.32
m_8	0.29	0.31	0.31	0.30	0.29	**0.27**	**0.27**	0.30	0.29
m_9	0.26	0.29	0.28	0.28	0.26	**0.25**	**0.25**	0.28	0.26

Table 4.29: Prediction Accuracy: Smaller Coefficients, Uniform Prior, Missing Predictor

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_{10}	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65
m_{11}	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71
m_{12}	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76

4.4 Simulated Data: Summary

The priors on g can be divided into two groups that have similar behavior: Group 1 being the n/2, Beta Prime, and Robust priors. Group 2 the Uniform, Hyper-g, Jeffery's, and Local EB priors.

From a model selection standpoint, the Group 1 priors on g outperform the Group 2 priors. The prediction results are nearly identical under all choices of g, but the Group 1 priors do generally slightly outperform Group 2. For what its worth, the Group 1 priors tend to result in a more concentrated posterior distribution across the model space, while the Group 2 priors tend to result in a flatter posterior distribution. So, if a flatter posterior distribution is desired, it may be appropriate to use a prior from Group 2. Otherwise, I would recommend using one of the Group 1 priors on g. In addition, BIC significantly outperforms AIC.

Among the two model space priors, the better choice is less clear. The Beta-Binomial(1,1) prior performs better when the correct model is a small or large percentage of the total potential predictors while the Uniform model space prior performs better when the correct model is around half of the potential predictors. So, the model space prior choice should depend on the a priori belief of the percentage

Table 4.30: Prediction Accuracy: Smaller Coefficients, Beta-Binomial(1,1) Prior, Missing Predictor

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_{10}	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65
m_{11}	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71
m_{12}	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76

Table 4.31: Mean Root Error: Smaller Coefficients, Uniform Prior, Missing Predictor

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_{10}	**0.50**	0.51	0.51	0.51	**0.50**	**0.50**	**0.50**	0.51	**0.50**
m_{11}	0.45	0.46	0.46	0.46	0.45	**0.44**	**0.44**	0.46	0.45
m_{12}	0.43	0.45	0.45	0.44	0.43	**0.42**	**0.42**	0.44	0.43

of the potential predictors that should be included in the resulting models.

Table 4.32: Mean Root Error: Smaller Coefficients, Beta-Binomial(1,1) Prior, Missing Predictor

	n/2	Uniform	Hyp-g	Jeffreys	Beta	AIC	BIC	EB	Robust
m_{10}	0.51	0.51	0.51	0.51	0.51	**0.50**	**0.50**	0.51	**0.50**
m_{11}	0.45	0.47	0.46	0.46	0.45	**0.44**	**0.44**	0.46	0.45
m_{12}	0.43	0.45	0.44	0.44	0.43	**0.42**	**0.42**	0.44	0.43

5

Application to Mortgage Default

5.1 Data and Models

The data is a 40,000 loan subset of fully amortized 30-year single-family fixed rate mortgages acquired by Freddie Mac. The selected loans originated in 2000 and contain loan payments through December 2012. 20,000 of these loans were used for training and 20,000 for testing. Data to be used was selected at random with the exception that manufactured housing was not included in this analysis. Manufactured housing has a significantly higher default rate and many of the predictors are in ranges outside of those under loans for more traditional housing types. For these reasons, I believe manufactured housing requires its own analysis, which will not be discussed here.

In the following analysis, default is defined as one of the following events:

1. 180 day delinquency
2. Deed-in-lieu of foreclosure prior to 180 day delinquency
3. Real Estate Owned (REO) Acquisition prior to 180 day delinquency

4. Short sale prior to 180 day delinquency

The large majority of defaults are simply 180 day delinquency and the other default events are comparatively rare. All predictors are the values at loan origination. The potential predictors to be considered are:

1. Credit Score: FICO credit score.

2. Loan-to-Value Ratio (LTV): For a standard loan, the ratio of original mortgage loan amount to the minimum of the property's appraised value and its purchase price. In the case of refinance, the ratio of the original mortgage loan amount to the appraised value.

3. Occupancy Status: Whether or not the property is the primary home for the purchaser (as opposed to vacation home, investment property, etc).

4. Unpaid Principle Balance (UPB): Original balance of loan.

5. Interest Rate: Original interest rate of loan.

6. Property Type: Whether or not the property is fee simple ownership. The vast majority of properties that are not fee simple ownership in the dataset are condos and planned unit development.

7. Loan Purpose: Whether loan is for purchase or refinance.

8. Number of Borrowers: One or more than one borrower.

In order to determine if polynomial terms should be considered, a generalized additive model was fit to the full model. Based on this analysis, 2nd and 3rd degree polynomials will be considered for credit score. Considering polynomial terms on the other predictors was deemed unnecessary. Through preliminary data analyses, I also decided to consider interaction terms between loan purpose and interest rate as well as 3rd degree and lower polynomials on credit score and number of borrowers.

For the models involving polynomials, models with higher order terms will only be considered if all lower order terms are included (ie a cubic term will only be considered for models that include both a linear and quadratic term). Models with interaction terms will only be considered if both predictors are included on their own as well.

Two priors will be considered on the model space. These are the Uniform prior and a prior distribution proportional to the Beta-Binomial(1,1) prior (which I will refer to as the Beta-Binomial (1,1) prior). A proportional version of the Beta-Binomial(1,1) prior needs to be used since not all combinations of predictors will be considered due to the previously mentioned restriction on polynomial terms and interaction terms.

I will look at prediction results under Bayesian Model Averaging under all of the described priors for g in Section 3.3. Specifically, these are the n/2, Uniform, Hyper-g, Jeffrey's, Beta Prime, Local EB, and Robust priors. I will also get results for AIC and BIC. For prediction under AIC and BIC, I will just use the top model chosen under the criteria.

5.2 Mortgage Data Results: Model Selection

Models will be referred to by model number from Table 5.1, that shows which predictors are included in each model.

Table 5.2 describes the posterior distribution across the five models with the highest posterior probabilities for the Uniform model space prior under each method except for AIC and BIC. For AIC and BIC, the table shows the selected model under the criteria. In the table, the top number refers to the model number and the bottom number is the posterior probability of the given model.

Table 5.2 shows that under the Uniform model space prior, all priors for g except Beta Prime and Robust choose model 7 as the model with the highest posterior likelihood. Beta Prime and Robust choose model 5. AIC and BIC choose models 14 and 5 respectively. The Uniform, Hyper g, Jeffrey's, and Local EB priors, choose the same models in the same order. The Beta Prime and Robust priors rank the same models in the same order as well. The n/2 prior is the same as the Beta Prime and Robust priors except that it switches the ordering of the top two models.

Model 7 includes all non-polynomial, non-interaction terms except occupancy status. It uses a 2nd degree polynomial for credit score, the linear number of borrowers:credit score interaction term, and does not include the loan purpose:interest rate interaction term. Model 5 is identical to Model 7 except it does not include the number of borrowers:credit score interaction term. Model 8 is the same as Model 5 except that it includes the loan purpose:interest rate interaction term. The model chosen under AIC, model 14, is Model 8 except that is uses a cubic polynomial on credit score.

Under the Uniform model space prior, none of the models in the top five highest posterior likelihoods under any prior contains occupancy status, the number of borrowers:credit score^2 interaction term, or the number of borrowers:credit score^3 interaction term. All models in the table contain the predictors credit score, LTV, UPB, interest rate, property type, and number of borrowers.

As shown in Table 5.2, not only do the Uniform, Hyper-g, Jeffrey's, and Local EB methods rank models identically across the top five, but the posterior likelihoods of each of these top five models are nearly identical as well. Each of these priors puts about 45% of the posterior distribution in the two most likely models and a little over 70% in the top five models. The n/2 prior is concentrated most heavily at the top two with over 80% of the posterior in the top two models and about 97% in the top five models. The Beta Prime and Robust posteriors are also significantly more concentrated towards the top with a little over 70% of the posterior in the top two models and about 90% in the top five. Although the Beta Prime and Robust priors have very similar results to each other, there is more differentiation in the posterior than that of the other group.

Table 5.3 shows the marginal inclusion probabilities (MIPs) under the Uniform model space prior. Credit Score, LTV, UPB, Interest Rate, Property Type, Loan Purpose, and Number of Borrowers have a MIP of about 1 for all choices of g prior. All priors also agree that there is little effect of occupancy status with an MIP of 0.12 as its highest value. The quadratic credit score term has a MIP of 1 under the Uniform, Hyper-g, Jeffrey's, and Local EB priors while it has an MIP of 0.8-0.9 under the others. No prior has a large MIP for the cubic credit score term with the n/2, Beta Prime, and Robust priors having a MIP of about 0.05 for it and the rest around 0.3. For the number of borrowers:credit score polynomial interactions, only the linear term has significant MIP while the quadratic term having an MIP of 0.13 at its highest and all priors giving the cubic term an MIP of around zero. The loan purpose:interest rate interaction is around 0.35 under the Uniform, Hyper-g, Jeffrey's, and Local EB prior while it is around zero on the others.

Table 5.4 describes the posterior distribution under the Beta-Bincomial(1,1) prior. When switching from the Uniform model space prior to the Beta-Binomial(1,1), the top model under n/2, Jeffrey's, and Robust priors stay the same. The top model under the Uniform, Hyper-g, and Local EB priors switch from Model 7 to Model 32, a model not in the top five under the Uniform model space prior. The Beta Prime prior switched its top and 2nd model from that under the Uniform prior. The Uniform, Hyper-g, Jeffrey's and Local EB priors no longer rank the top five models identically, but they do have the same unordered top five. Nor do the Beta Prime and Robust priors, although it is the same except for a switched top and 2nd ranked model. The rankings under the n/2 prior are the only ones that haven't changed at all.

The Model 32 that has become the top model under a few of the priors is the full model. This makes sense as the Beta-Binomial(1,1) prior puts significant probability on the top model a priori. Another model that was not on the Uniform model space prior version of this table is Model 18 which is the full model not including Occupancy Status. In general, using the Beta-Binomial(1,1) prior has resulted in generally larger models chosen than under the Uniform model space prior.

As shown in Table 5.4, the Beta-Binomial(1,1) prior resulted in a much flatter posterior distribution across the model space under the Uniform, Hyper-g, Jeffrey's, and Local EB priors. Under the Uniform model space prior, the top model for these priors had a posterior likelihood of just under 30% and now they are just over 10%. In fact, there is only a 2%-3% difference between the likelihood of the top model and that under the 5th model under these priors. The total posterior probability across the top five models decreased from just over 70% to just over 50%. Across the top two models the posterior probability dropped from about 45% to just over 20%. The corresponding posterior probabilities under the other priors barely changed.

Table 5.5 shows the resulting MIPs under the Beta-Binomial(1,1) prior. Just as under the Uniform model space prior, the linear Credit Score term, LTV, UPB, Interest Rate, Property Type, Loan Purpose, and Number of Borrowers have a MIP of about 1 for all choices of g prior. The results under the n/2, Beta Prime and Robust priors are essentially the same as under the Uniform prior, but with slightly higher MIPs for those predictors with smaller MIPs under the Uniform model space prior. Under the other priors, there was substantial increases in MIP for the cubic Credit Score term, Occupancy Status, the Number of Borrowers:quadratic and cubic Credit Score interaction terms, and Loan Purpose:Interest Rate interaction term.

5.3 Mortgage Data Results: Prediction

Tables 5.6-5.9 show the prediction results for the 20,000 loan test set. As default is a relatively rare event, the predicted default rate is never $> 50\%$. So, instead of predicting default, I grouped the test data by predicted default rate and looked at the actual default rates in each group. In these tables, the row groupings correspond to the predicted default rates. In tables 5.6 and 5.8, the values in the table correspond to the actual default rates for loans with predicted default rates in the given group. In tables 5.7 and 5.9, the values correspond to the total loans with default rates predicted to be in the given group.

Tables 5.6-5.9 show that with this large data set, prior choice does not have a large impact. The only criteria that fails to produce an actual default rate within the predicted range is the 5%-10% group under AIC.

Over 11,000 of the 20,000 loans have predicted default rates of less than 1% and the actual default rates of these loans are about 0.3%. Only about 2,500 of 20,000 loans have predicted default rates of above 5% and about 900 loans have predicted default rates above 10%.

The choice of prior on g does not have a large affect on the predictions, although there is still a small difference in results between priors. Even AIC and BIC produce similar results to those under Bayesian Model Averaging.

The choice of prior on the model space only has a small effect as well. When changing the model space prior, the counts of default predictions never change by more than 18 in the $< 1\%$ bucket, 38 in the $1\% - 5\%$ bucket, 27 in the $5\% - 10\%$ bucket, 4 in the $10\% - 15\%$, and 7 in the $> 15\%$ bucket.

Tables 5.10 and 5.11 show the posterior means of the coefficients on each of the predictors. Note that Credit Score and Number of Borrowers have not been included in these tables. To show the overall effect of Credit Score and Number of

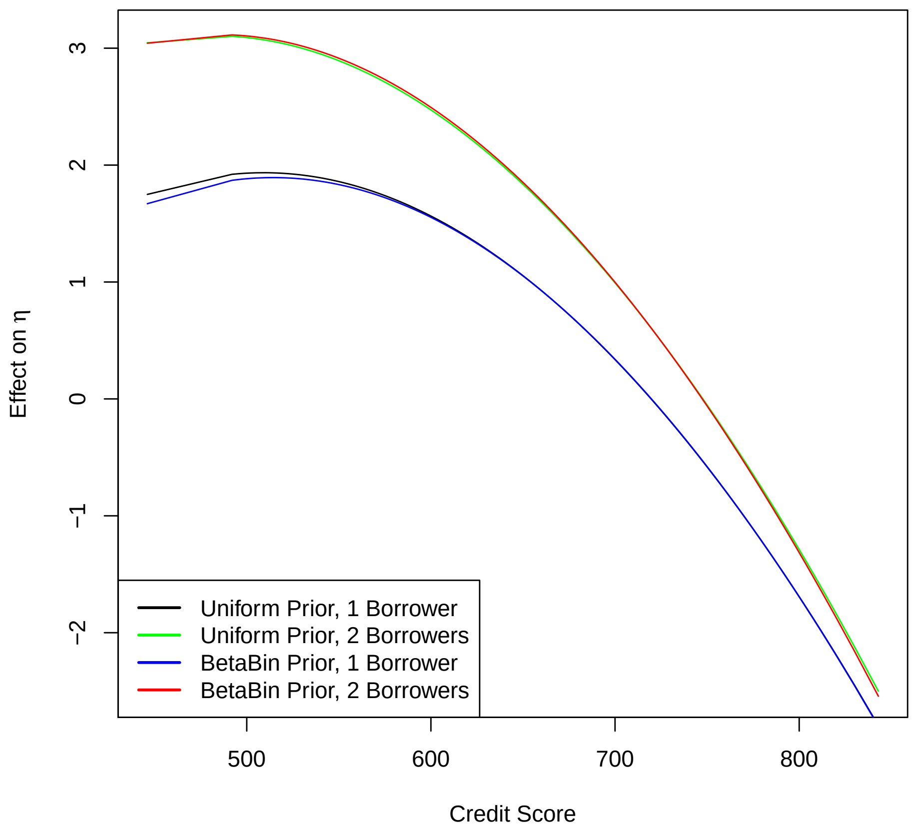

FIGURE 5.1: Effect of Credit Score and Number of Borrowers on η: n/2 Prior

Borrowers, see Figures 5.1 and 5.2. These Figures show the effect of Credit Score on the predicted η's by Number of Borrowers and model space prior. As Uniform, Hyper-g, Jeffrey's, and Local EB have nearly identical coefficients, I showed the Uniform Prior as representative of these four priors. I showed the n/2 prior as representative of the n/2, Beta Prime, and Robust priors as these priors have nearly identical coefficients on the given predictors as well.

Tables 5.10 and 5.11 show that Occupancy Status has essentially no effect as does Unpaid Principle Balance. Increased LTV, leads to a higher probability of default as does the loan's Interest Rate. A Fee Simple property has a little higher probability of

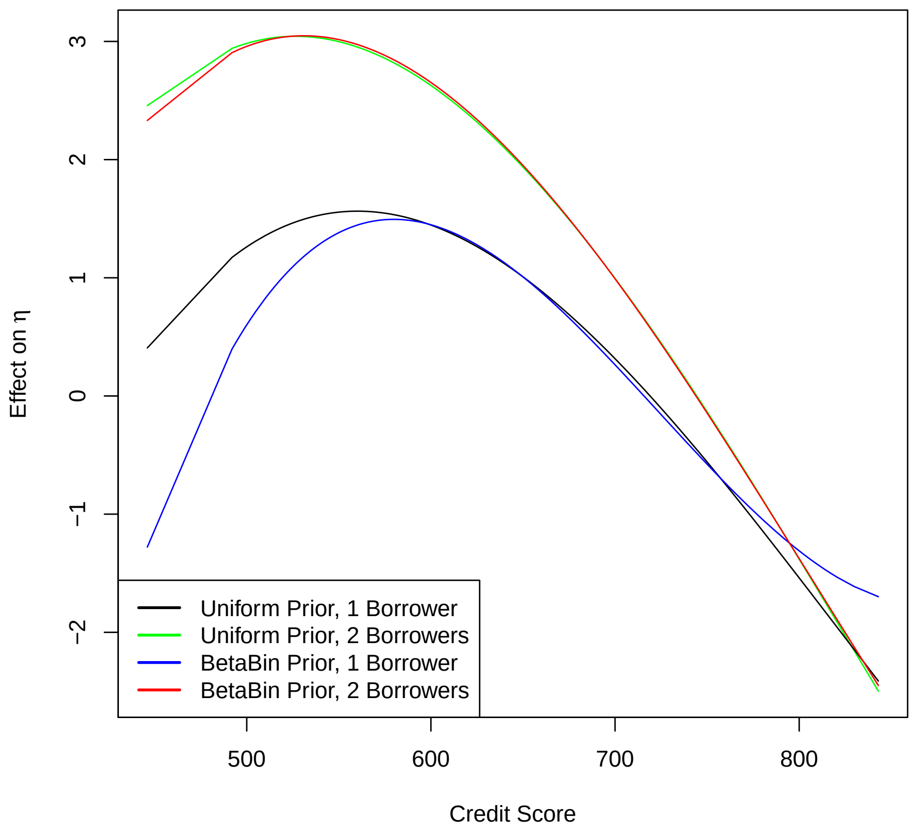

FIGURE 5.2: Effect of Credit Score and Number of Borrowers on η: Uniform Prior

default than condos and planned unit development. Loans for refinance as opposed to purchase also have a higher probability of default. The only coefficient that changes by a large amount in these tables by switching model space priors is the Loan Purpose coefficient which increases a bit under the Uniform, Hyper-g, Jeffrey's, and Local EB priors when switching from the Uniform to Beta-Binomial(1,1) prior. There is also a little bit of a decrease in the Loan Purpose:Interest Rate interaction term under these priors after making the change.

Figure 5.1 shows that under the n/2, Beta Prime, and Robust priors, similar to the results shown in Tables 5.10 and 5.11, model space prior choice does not affect results much. It shows that having more than one borrower increases probability of default, but this difference gets smaller and smaller as Credit Score increases. For the other priors, Figure 5.2 shows that for more than one borrower, the effect of Credit Score and Number of Borrowers is similar to that as under the priors in Figure 5.1. When there is only one borrower, the results are quite a bit different and the shape of the effect is a little surprising. This is likely due to the small amount of data with Credit Score less than 600. For one borrower, the model space priors have the same affect in areas with the bulk of the data, but diverge for Credit Scores under 600 and over 800 which are the areas with small amounts of data.

Table 5.1: Included Predictors by Model

	CredScore	OccStatus	LTV	UPB	IntRate	PropType	LoanPurp	NumBorr	CredScore^2	CredScore^3	NumBor*CredScore	NumBor*CredScore^2	NumBor*CredScore^3	LoanPurp*IntRate
1	1	0	1	1	1	0	1	1	1	0	0	0	0	0
2	1	0	1	1	1	1	1	1	0	0	0	0	0	0
3	1	0	1	1	1	1	1	1	0	0	0	0	0	1
4	1	0	1	1	1	1	1	1	0	0	1	0	0	0
5	1	0	1	1	1	1	1	1	1	0	0	0	0	0
6	1	0	1	1	1	1	1	1	1	0	0	0	0	1
7	1	0	1	1	1	1	1	1	1	0	1	0	0	0
8	1	0	1	1	1	1	1	1	1	0	1	0	0	1
9	1	0	1	1	1	1	1	1	1	0	1	1	0	0
10	1	0	1	1	1	1	1	1	1	0	1	1	0	1
11	1	0	1	1	1	1	1	1	1	1	0	0	0	0
12	1	0	1	1	1	1	1	1	1	1	0	0	0	1
13	1	0	1	1	1	1	1	1	1	1	1	0	0	0
14	1	0	1	1	1	1	1	1	1	1	1	0	0	1
15	1	0	1	1	1	1	1	1	1	1	1	1	0	0
16	1	0	1	1	1	1	1	1	1	1	1	1	0	1
17	1	0	1	1	1	1	1	1	1	1	1	1	1	0
18	1	0	1	1	1	1	1	1	1	1	1	1	1	1
19	1	1	1	1	1	1	1	1	1	0	0	0	0	0
20	1	1	1	1	1	1	1	1	1	0	0	0	0	1
21	1	1	1	1	1	1	1	1	1	0	1	0	0	0
22	1	1	1	1	1	1	1	1	1	0	1	0	0	1
23	1	1	1	1	1	1	1	1	1	0	1	1	0	0
24	1	1	1	1	1	1	1	1	1	0	1	1	0	1
25	1	1	1	1	1	1	1	1	1	1	0	0	0	0
26	1	1	1	1	1	1	1	1	1	1	0	0	0	1
27	1	1	1	1	1	1	1	1	1	1	1	0	0	0
28	1	1	1	1	1	1	1	1	1	1	1	0	0	1
29	1	1	1	1	1	1	1	1	1	1	1	1	0	0
30	1	1	1	1	1	1	1	1	1	1	1	1	0	1
31	1	1	1	1	1	1	1	1	1	1	1	1	1	0
32	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 5.2: Model Space Posterior: Uniform Prior

	n/2	Uniform	HypG	Jeffreys	Beta	AIC	BIC	EB	Robust
Top Model	7	7	7	7	5	14	5	7	5
Post Prob	0.43	0.29	0.30	0.31	0.39			0.30	0.44
2nd	5	8	8	8	7			8	7
Post Prob	0.37	0.17	0.17	0.16	0.33			0.16	0.27
3rd	2	13	13	13	2			13	2
Post Prob	0.08	0.14	0.14	0.14	0.12			0.14	0.17
4th	6	14	14	14	6			14	6
Post Prob	0.04	0.09	0.08	0.08	0.03			0.08	0.03
5th	11	5	5	5	11			5	11
Post Prob	0.04	0.04	0.05	0.05	0.03			0.05	0.02

Table 5.3: Marginal Inclusion Probabilities: Uniform Prior

	n/2	Uniform	HypG	Jeffreys	Beta	AIC	BIC	EB	Robust
CredScore	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
CredScore2	0.90	1.00	1.00	1.00	0.86	1.00	1.00	1.00	0.81
CredScore3	0.04	0.33	0.32	0.31	0.06	1.00	0.00	0.33	0.04
I(Occupied)	0.01	0.12	0.11	0.10	0.01	0.00	0.00	0.11	0.01
LTV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
UPB	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
IntRate	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
I(FeeSimple)	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.99
I(Refinance)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
I(2Borrowers)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
I(2Bor)*CScore	0.45	0.90	0.89	0.88	0.41	1.00	0.00	0.89	0.32
I(2Bor)*CScore2	0.00	0.13	0.12	0.11	0.01	0.00	0.00	0.12	0.00
I(2Bor)*CScore3	0.00	0.02	0.02	0.01	0.00	0.00	0.00	0.02	0.00
I(Ref)*IntRate	0.05	0.38	0.37	0.35	0.07	1.00	0.00	0.36	0.05

Table 5.4: Model Space Posterior: Beta-Binomial(1,1) Prior

	n/2	Uniform	HypG	Jeffreys	Beta	AIC	BIC	EB	Robust
Top Model	7	32	32	7	7	14	5	32	5
Post Prob	0.49	0.12	0.11	0.11	0.35			0.11	0.38
2nd	5	18	14	14	5			14	7
Post Prob	0.33	0.11	0.11	0.11	0.32			0.11	0.30
3rd	2	14	7	8	2			7	2
Post Prob	0.07	0.10	0.10	0.10	0.10			0.10	0.14
4th	6	7	18	32	8			18	8
Post Prob	0.05	0.09	0.10	0.10	0.05			0.10	0.03
5th	11	8	8	18	13			8	13
Post Prob	0.04	0.09	0.10	0.09	0.05			0.10	0.03

Table 5.5: Marginal Inclusion Probabilities: Beta-Binomial(1,1)

	n/2	Uniform	HypG	Jeffreys	Beta	AIC	BIC	EB	Robust
CredScore	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
CredScore2	0.91	1.00	1.00	1.00	0.89	1.00	1.00	1.00	0.83
CredScore3	0.04	0.67	0.65	0.63	0.09	1.00	0.00	0.65	0.06
I(Occupied)	0.01	0.34	0.32	0.30	0.02	0.00	0.00	0.32	0.01
LTV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
UPB	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
IntRate	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
I(FeeSimple)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
I(Refinance)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
I(2Borrowers)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
I(2Bor)*CScore	0.50	0.97	0.96	0.95	0.49	1.00	0.00	0.96	0.40
I(2Bor)*CScore2	0.00	0.48	0.45	0.43	0.01	0.00	0.00	0.46	0.01
I(2Bor)*CScore3	0.00	0.29	0.27	0.25	0.00	0.00	0.00	0.27	0.00
I(Ref)*IntRate	0.06	0.65	0.63	0.61	0.11	1.00	0.00	0.63	0.07

Table 5.6: Grouped Predicted vs. Actual Default Rates: Uniform Prior

	n/2	Uniform	HypG	Jeffreys	Beta	AIC	BIC	EB	Robust
<1%	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
1%-5%	0.023	0.023	0.023	0.023	0.023	0.025	0.022	0.023	0.023
5%-10%	0.053	0.052	0.052	0.052	0.053	0.046	0.057	0.052	0.053
10%-15%	0.117	0.113	0.113	0.113	0.113	0.111	0.112	0.113	0.114
>15%	0.186	0.191	0.190	0.189	0.189	0.189	0.188	0.190	0.192

Table 5.7: Total Predictions in Group: Uniform Prior

	n/2	Uniform	HypG	Jeffreys	Beta	AIC	BIC	EB	Robust
<1%	11203	11138	11144	11154	11199	11450	11234	11155	11175
1%-5%	6269	6264	6259	6252	6277	5987	6310	6250	6326
5%-10%	1593	1659	1657	1652	1591	1596	1552	1653	1577
10%-15%	521	542	540	540	520	539	490	541	510
>15%	414	397	400	402	413	428	414	401	412

Table 5.8: Grouped Predicted vs. Actual Default Rates: Beta-Binomial(1,1) Prior

	n/2	Uniform	HypG	Jeffreys	Beta	AIC	BIC	EB	Robust
<1%	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
1%-5%	0.023	0.023	0.023	0.023	0.023	0.025	0.022	0.023	0.023
5%-10%	0.052	0.051	0.051	0.051	0.053	0.046	0.057	0.051	0.053
10%-15%	0.116	0.114	0.114	0.115	0.115	0.111	0.112	0.114	0.115
>15%	0.187	0.186	0.184	0.184	0.185	0.189	0.188	0.184	0.189

Table 5.9: Total Predictions in Group: Beta-Binomial(1,1) Prior

	n/2	Uniform	HypG	Jeffreys	Beta	AIC	BIC	EB	Robust
<1%	11212	11128	11135	11154	11209	11450	11234	11156	11193
1%-5%	6248	6292	6283	6263	6255	5987	6310	6264	6288
5%-10%	1604	1632	1632	1635	1596	1596	1552	1630	1592
10%-15%	519	544	543	540	524	539	490	543	514
>15%	417	404	407	408	416	428	414	407	413

Table 5.10: Avg Posterior β's: Uniform Prior

	n/2	Uniform	HypG	Jeffreys	Beta	AIC	BIC	EB	Robust
Intercept	-2.94	-2.93	-2.93	-2.93	-2.94	-2.92	-2.95	-2.93	-2.94
I(Occupied)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LTV	0.05	0.04	0.04	0.04	0.05	0.05	0.05	0.04	0.05
UPB	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
IntRate	0.43	0.46	0.46	0.46	0.43	0.54	0.42	0.46	0.43
I(FeeSimple)	0.62	0.61	0.61	0.61	0.62	0.62	0.62	0.61	0.62
I(Refinance)	1.19	1.93	1.90	1.87	1.23	3.50	1.06	1.91	1.20
I(Ref)*IntRate	-0.01	-0.11	-0.10	-0.10	-0.02	-0.29	0.00	-0.10	-0.02

Note: See Figures 5.1 and 5.2 for effect of credit score and number of borrowers

Table 5.11: Avg Posterior β's: Beta-Binomial(1,1) Prior

	n/2	Uniform	HypG	Jeffreys	Beta	AIC	BIC	EB	Robust
Intercept	-2.94	-2.92	-2.92	-2.92	-2.94	-2.92	-2.95	-2.92	-2.94
I(Occupied)	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.00
LTV	0.05	0.04	0.04	0.04	0.05	0.05	0.05	0.04	0.05
UPB	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
IntRate	0.43	0.49	0.49	0.49	0.43	0.54	0.42	0.49	0.43
I(FeeSimple)	0.62	0.61	0.61	0.61	0.62	0.62	0.62	0.61	0.62
I(Refinance)	1.20	2.57	2.53	2.49	1.32	3.50	1.06	2.54	1.24
I(Ref)*IntRate	-0.02	-0.18	-0.18	-0.17	-0.03	-0.29	0.00	-0.18	-0.02

Note: See Figures 5.1 and 5.2 for effect of credit score and number of borrowers

6

Conclusion

In conclusion, the priors on g can be put into two groups where the the priors in each group behave similarly. Group 1 being the n/2, Beta Prime, and Robust priors. Group 2 is the Uniform, Hyper-g, Jeffrey's, and Local EB priors. The Group 1 priors tend to have more concentrated posterior distributions across the model space.

The simulated data study showed that in model selection the Group 1 priors outperform the Group 2 priors. For prediction, the Group 1 priors still outperformed Group 2, but the results were very close. For the large mortgage default data set, the Group 1 and Group 2 priors did choose slightly different models, but there was little difference in the predicted values under the different models chosen. In general, I would recommend using one of the Group 1 priors for g.

For the two model space priors, there is no clear better choice. The Beta-Binomial(1,1) prior was more successful under the small and full models in the simulated data while the Uniform prior was more successful under the middle sized model. This is more of a situational choice, but using the Uniform prior when p is very large under the full model, will likely result in a middle sized model being selected as the prior likelihoods of the very small and very large models will be extremely small.

Bibliography

[1] Bayarri, M. J., Berger, J. O., Forte, A., and Garcia-Donato, G. (2012), *"Criteria for Bayesian model choice with application to variable selection"*, The Annals of Statistics, 40, 1550-1577.

[2] Celeux, G., Anbari, M. E., Marin, J., and Robert, C. (2012), *"Regularization in Regression: Comparing Bayesian and Frequentist Methods in a Poorly Informative Situation"*, Bayesian Analysis, 7, 477-502.

[3] Feldkircher, M., Zeugner, S., (2009), *"Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging"*, IMF Working Papers, 1-39.

[4] Fernandez, C., Ley, E., Steel, M. (2001), *"Model Uncertainty in Cross-Country Growth Regressions"*, Journal of Applied Econometrics, 16, 563-576.

[5] Gordy, M.B. (1998) *"Computationally Convenient Distributional Assumptions for Common Value Auctions"*, Computational Economics, 12.

[6] Ley, E., Steel, M. (2009), *"On the Effect of Prior Assumptions in Bayesian Model Averaging With Applications to Growth Regression"*, Journal of Applied Econometrics, 24, 651-674.

[7] Ley, E., Steel, M., (2012), *"Mixtures of g-priors for Bayesian Model Averaging with Economic Applications"*, Journal of Econometrics, 171, 251-266.

[8] Li, Y. (2013), *"Bayesian Hierarchal Models for Model Choice"*.

[9] Hansen, M. H. and Yu, B. (2001), *"Model Selection and the Principle of Minimum Description Length"*.

[10] Liang, F., Paulo, R., Molina, G., Clyde, M., and Berger, J. (2008), *"Mixtures of g Priors for Bayesian Variable Selection"*, Journal of the American Statistical Association, 103, 410-423.

[11] Maruyama, Y. and George, E. I. (2011), *"Fully Bayes Factors with a Generalized g-prior"*, The Annals of Statistics, 39, 2740-2765.

[12] Wang, X. and George, E.I. (2007), *"Adaptive Bayesian Criteria in Variable Selection for Generalized Linear Models"*, Statistics Sinica, 17, 667-690.